Digital Enlightenment Yearbook 2013
M. Hildebrandt et al. (Eds.)
IOS Press, 2013
© 2013 The authors.
doi:10.3233/978-1-61499-295-0-270

Personal Data Management – A Structured Discussion

Jacques BUS[a,1] and M.-H. Carolyn NGUYEN[b]
[a]*Digital Enlightenment Forum*
[b]*Microsoft*

Abstract. In the coming decade, a seamless integration of our on- and offline lives will be necessary for a sustainable digital society. This requires urgent multidisciplinary debate on the collection, control and use of personal data in society. This paper proposes a framework that can be used to shape this dialogue. It is based on a neutral and consistent terminology that may support a constructive and fruitful debate, avoiding terms that have too often led to controversy and confusion. We have attempted to position in this structure state-of-the-art technology developments, including context-awareness, user-centred data management, trust networks and personal data ecosystems. It demonstrates clear relations between ongoing work in various groups and hence an urgent need for cooperation to achieve common goals. We argue that such cooperation can lead to the emergence of a personal data ecosystem that may truly support a sustainable digital society.

Keywords. Personal data management, identity, privacy, context-awareness, trust networks, user-centricity, personal data ecosystem

Introduction

In the coming decade, the seamless integration of our on- and offline lives will become an increasingly important issue facing society. It will require that individuals be able to manage how information related to them is used, whether directly or through other means, in a way that meets with their preferences, contexts, and values, all within existing social and legal boundaries. This will be a departure from current practice, where users would give automatic consent to their data being processed, often without any true understanding of the full context in which the data will be used. Understanding of the impact of this change is crucial, and broad societal discussion is necessary to ensure sustainable social and economic development. This can be challenging, especially in a world of big data, where individuals may not be aware of the majority of the data that exists – or can reveal information – about them. Enabling individuals to exert some control over how these data are used will be an important aspect of an overall solution.

Big data poses other challenges to existing approaches in data privacy, pointing to the need for dialogues among stakeholders on a policy framework that can help create a sustainable data ecosystem that will drive new business models and innovation while also strongly protecting individual rights. The World Economic Forum highlighted some of these issues in its report [1]; the Digital Enlightenment Forum, with this Year-

[1] Corresponding Author.

book and its 2013 Forum, is bringing together experts to develop a coherent framework and consistent policy recommendations to help address them.

This paper aims to facilitate and further these discussions by presenting a possible framework that can be used to shape the dialogue, and select technological development and issues that should be taken into consideration. We first try to develop language and terminology that might help to establish some structure and make progress in the discussions. Words like identity, privacy, personal data, trust, context and many more are used by different people in different ways, leading to disagreement and confusion. Moreover, priorities are looked at differently from different disciplines. In a second step, we describe ongoing technologies and developments that can contribute to a sustainable structure of personal data management.

1. Privacy, Personal Data Management, and Identity

Privacy, personal data management, and *identity* are controversial issues in policy development for the digital society. Unfortunately, the debate suffers from a persistently confusing use of terminology by various stakeholders, and any progress will require some mutual agreement around these concepts. We start, therefore, with some discussion on the nuances of the terminology, before integrating them into a layered model that can be used to address issues related to use of personal data in the digital society that would take into consideration individual preferences and context.

First, although *privacy* and *personal data management* (*PDM*) are rightly seen as overlapping, they are not the same. Some perceive privacy to be broader than PDM, as it also includes physical privacy, "the right to be left alone", the dignity of the person and other aspects of social behaviour. For others, privacy is considered narrower than PDM, as the latter includes the management of the complete lifecycle of data that may relate to a person (e.g. attributes of relevance to an individual, context for sharing personal data, and decisions on fair value exchange). For them, privacy is just one aspect of a holistic view of PDM. In addition, although privacy often connotes withholding personal information to ensure the "right to be left alone", PDM connotes managing the flow, access, and use of data to ensure that it can deliver perceived value to the individual. Depending on the individual, this value can be at the personal, social, economic or global level.

In the context of the consequences of digitisation, it therefore seems more relevant to discuss PDM than privacy. Thinking about PDM, it must be noted that what is considered personal is a complex concept, and is also interpreted very differently in different contexts, cultures, social and political environments (see, e.g. James Q. Whitman [2] and the report from the International Institute of Communications [3]). This means that if we want to consider personal data management as a universal concept, its definition should not be dependent on the specific social/cultural environment in which we apply it. Furthermore, an operational definition must integrate the contexts of data use, including social and cultural norms. Such a definition would also need to reflect the fact that preferences and norms can change over time.

Another source of confusion is the term *identity*. In the governmental context, identity is often interpreted as the set of data (attributes) required to uniquely and authoritatively identify a person, which is then often related to a unique identifier (e.g. a citizen registration number). The proposed EC eID regulation, for example, demands assurances from Member States that their "notified eID systems" for authentication lead to a single, unique citizen. However, modern technology for authentication based on attributes aims at data minimisation, as described in the OECD guidelines for privacy, and does not require a unique identifier – only that the attributes relevant to the given transaction be verified to allow access. When applied cross-border, standards would thus be needed to ensure that such assurances, when used to access a public service (including passing a national border), could also satisfy the uniqueness requirement mandated at a national level.

In many situations, it is sufficient and desirable to claim only one or a few attributes – often without name and address – to obtain access to a public or private service (e.g. the age of a person purchasing liquor or the assurance that a credit card is valid) or to communicate with others. Moreover, social behaviour in the physical world shows us that people present themselves differently (sharing different data, emphasising particular personal details) in different environments (family, work, sports club, friends, public place, etc.). This "multiple identity" behaviour, including the ability to be anonymous or pseudonymous, must be supported in the digital world if the concept of a digital society is to be acceptable and trusted.

For clarity of the discussion that we want to start here, we will avoid the word "*privacy*" and use the term "*context-aware PDM*" and "*identity*" as described below:

> Context-aware PDM (CPDM) enables an individual to control the access and use of her personal data in a way that gives her sufficient autonomy to determine, maintain, and develop her identity as an individual, including the choice of attributes of her identity to present depending on the context of the transactions.

CPDM will enable consideration of constraints relevant to personal preferences, cultural, social, and legal norms. Trustworthy data practices are foundational to enabling CPDM and hence issues of security, integrity, and protection of personal data, as well as data governance, auditability and other aspects related to trust must be recognised and addressed.

This description contains undefined elements (such as *sufficient* and *context*). These depend on the implementing environment (e.g. purpose of data use, personal value derived, social norms, cultural rules, etc.) and make it possible to discuss the concept of CPDM at a universal level. One person in a specific context may make personal choices different from another person in the same context. The same person may also make different choices regarding the same data in different contexts. This does not change the essence of the right to privacy, but rather reinforces it as including the right to such choices, and hence the right to a certain socially and culturally acceptable autonomy.

Also note that, although we discuss the concept of data use in the context of a *transaction*, it is not necessary that the individual is actually *engaged* in the transaction. For example, she could agree that her health data may be used in medical research; in this case, the research is the "transaction". Regardless of her level of engagement in the transaction, the use of her data must still meet the constraints defined above.

Although the description of CPDM is applicable for various jurisdictions, it does not explicitly address cross-border or cross-jurisdictional aspects. In some ways, the description assumes these.

For the definition of CPDM, we have chosen to emphasise one's choice to develop personal identity and manage personal data within given contexts, as we feel that this is a critical enabler for integrating our on- and offline lives.

2. User Control of Personal Data

Let us first come back to the concept of *personal data*. In the EC proposal for a Data Protection Regulation, it is defined as follows:

> "Personal Data" means any information relating to a data subject, being an identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by a data controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person [4].

In the US, the term *Personally Identifiable Information* (*PII*) is used, which is essentially the same.

Until recently, a commonly accepted approach for preserving the anonymity of personal records and providing some minimal privacy protection was to remove any "personal data" such as names and social security numbers from such records before storing them in large databases or sharing them for public use. However, recent works have demonstrated that by aggregating multiple databases, the records can be easily "re-identified" or "de-anonymised" [5]. An early case that is quite well-known involved a Netflix contest, which ran from October 2006 to Sept 2009, where $1 million was to be awarded to the team that could best improve the company's movie-recommendation algorithm [6]. Contestants were given access to members' movie ratings to develop an algorithm that could accurately predict movies those same users would enjoy. Two massive data sets were released: 100 million movie ratings from 480,000 customers, along with the date of the rating, subscriber unique ID, and the movie info – all data that that were considered to be anonymous and could not lead to identification, and therefore would not traditionally be considered personal data. Nevertheless, Netflix was sued for breach of consumer privacy after researchers at the University of Texas proved that the movie ratings could be de-anonymised and traced to identifiable individuals [7].

For big data, where diverse data sets are commingled to derive new insights, these research findings make it less clear how personal data should be defined. If taken to the extreme, they may lead one to conclude that any data may potentially be classified as personal data. Since many data protection frameworks today limit inappropriate use of personal data by restricting its processing, and do not differentiate among the different types of personal data, the combination of these two approaches would have a serious impact on enabling the value of big data to be realised.

In the remainder of this paper, we use the term *personal data* (*PD*) generically to define any data that can be related to an individual, since any such data would have the ability to impact an individual's identity. However, we believe that an alternative policy approach to CPDM is needed to ensure that the value potential of big data can be realised – one that is based on data use rather than data collection or processing. This is a driver in our development of this paper and is addressed further below.

In discussing CPDM in an increasingly data-driven world, it is relevant to distinguish:

- *Actively collected PD* provided by the user as part of a transaction (e.g. address, credit-card information);
- *Passively generated PD*, which include data that are either (a) collected without user awareness as part of an active transaction (e.g. location information

while completing a purchase); (b) collected without user awareness and users are also unaware of any ongoing transaction (e.g. video camera capture as individuals walk through a train station); (c) inferred data resulting from analytics of data that can be aggregated about the user (e.g. credit-report rating).[2]

For actively collected PD, users can control the flow and use of their data by giving or withholding consent at the point of data collection – this is the notice/consent model currently in use today. As indicated above, for most users, consent is given without any true understanding of the full context in which the data will be used, or of the potential benefits obtained beyond the immediate service.

With big data, the majority of data collected will increasingly be passively generated. For these data, it is neither practical, nor in some cases possible, for users to exercise explicit control. The aforementioned World Economic Forum report discussed the need for a new approach to managing PD in a decentralised data-driven ecosystem to balance the socio-economic value that can be unleashed by what we term CPDM. It also discussed the shift to a use-based approach for these data that takes personal context into consideration along with new ways to engage individuals. Underlying these changes are a set of updated principles for CPDM, complemented with enforceable codes of conduct and technology that can facilitate the implementation of trustworthy data practices and contextually aware data usage in such an ecosystem.

In addition to controlling the use of data, another aspect of CPDM would seem to require that there be protection against unauthorised access of data. This includes preventing entities that cannot adequately verify their identities from accessing and using the data, or preventing a breach of the data while it is at rest (e.g. in storage) or in transit. These measures are necessary, but not sufficient, elements to ensure fundamental trustworthiness of the system, regardless of whether the personal data are actively collected or passively generated.

When the data are stored in users' personal storage devices, the user would be in control and accountable for keeping the data secured using technologies such as data encryption. When the data are stored in a cloud, or databases managed/owned by third parties, the appropriate service providers are accountable for keeping the data secured and conditions would be agreed for enforcement. If the service provider is providing a secured personal data storage service, it may also agree with the user to keep the data encrypted and provide them with a private key. Of course, keeping stored data secured will not mean secured data processing. Some possible ways of improving the secure management of data in transit are discussed later.

3. A Reference Model for Personal Data Management

The issues that need to be addressed in achieving CPDM can be structured into three "layers":

[2] The chapter by Nguyen et al. in this volume discusses how, in the world of big data, the number of devices, sensors, and other objects that can collect data will vastly exceed the number of people on the planet, and that the majority of big data will be collected passively and automatically, through machine-to-machine transactions, without user involvement. The choice of terminology used in that chapter, as well as here, *actively collected PD* and *passively generated PD*, emphasises these facts and reinforces the need for new approaches towards policy development. The World Economic Forum uses the terms *volunteered*, *observed*, and *inferred* data to refer to same in [1].

- Infrastructure
- Data management
- User interaction.

This model can also provide a useful frame of reference to debate questions regarding future developments on technological, economic, social, and policy issues relevant to PDM. These elements must be balanced and co-exist at every level in the model below.

The essence of the debate on management of personal data and citizenship in the digital society is driven by how the elements of the three layers in this model should be specified and implemented; what are the guiding principles for enabling them to provide trustworthy data practices; and how they can balance the needs of individuals, industry, and regulators.

3.1. Infrastructure

The *infrastructure* layer contains the services and applications required to assure the integrity and security of the data, both while in transit and at rest, including supporting the appropriate protocols, service bindings, etc. Processes facilitated by the infrastructure should be trustworthy. This will entail an architecture that has been designed with security and data protection in mind (Security and DP by Design). Technical considerations include encryption and using techniques such as differential privacy to protect against unintended personal data breaches and system attacks.[3] It should facilitate logging, monitoring and random control of compliance with the basic rules agreed for the infrastructure and the processes facilitated by it. Services provided by these elements should be available through Application Programming Interfaces (API) as part of the platform.

Authentication (verifying identity or claims[4] in general), and the services provided by *Trusted Identity* (*or Claim*) *Providers* (TIP) at various levels of assurance, including anonymous or pseudonymous claims, are essential elements of this layer to enable a foundation for trustworthy practices.

Trustworthiness and acceptability of the infrastructure may be achieved through market mechanisms (reputation, brands, price), through regulation, certification, control, and enforcement, or through other more direct forms of governance directly supervising (parts of) the infrastructure. The choices here depend on the culture and the political environment. In western societies, governance would usually include sufficient checks and balances that aim at separation of concerns and avoidance of conflicts of interest.

3.2. Data Management

The *data management* layer includes the elements (apps, services, etc.) required to enable individuals and service providers to effectively control the flow and use of per-

[3] Differential privacy refers to techniques that would enable analysts to extract useful insights about the population as a whole from a database containing personal information, while at the same time protecting the individuals from being identified in the sample [19]. These techniques work by introducing small distortions into the results in a way that would not invalidate the results extracted, but would protect individual privacy. As such, they prevent unwanted re-identification of the personal data contained in the database.

[4] Claims are defined here as assertions about one or more attributes for a given persona.

sonal data based on specified use permissions and policies. Elements of this layer include mechanisms that would enable individuals to specify data use permissions, service providers to communicate data use policies, data policy enforcement services, and capabilities to audit data use compliancy. Trustworthy data practices are critical at this layer to enhance user trust in the overall data ecosystem. One method of improving trustworthiness would be a contract between a data controller and the user that the controller is accountable to abide by the use permissions specified by the user. Technology such as an interoperable metadata-based architecture that (logically) binds permissions and policies to the data can provide the means for facilitating the enforcement of these contracts (see the chapter by Nguyen, et al. in this volume).

Service providers would leverage services offered by the infrastructure and use the API management system(s) provided. For example, TIP can provide the claims needed to authenticate an entity and enable it to access and utilise a given piece of data based on the data policy specified in the metadata.

Services that can be offered based on the elements provided at this layer include user reputation management services, trusted social networks information sharing, consumer-trusted credit ratings, services to facilitate consent management, advertising management, auditing, data access requests, and many more we cannot yet envisage.

3.3. User Interaction and Context

The *user interaction* layer includes the elements that enable end users to have a meaningful interaction with service providers regarding the permissions and policies associated with the use of their personal data. Clearly, what is meaningful and intuitive in one culture or jurisdiction might not be so elsewhere. And it goes further than simply implementing local legal compliance and cultural norms. The user experience and interaction models to be developed would need to take into consideration the users' mental models on personal data, reflecting potential differences in shifting personal preferences, as well as social and cultural norms. And it should allow the emotional and autonomous choices of the user in the given context.

Today, data use is normally expressed as a binary decision. Either the user agrees to enable use of the data collected or not. In the physical world, data use is highly nuanced and contextual. The same data that are considered sensitive by one person may not be considered so by another. Even for data that are considered sensitive, their uses are highly dependent on the context and the assumed identity or persona at a given time. Consider location data as an example. Some users may never want their location data to be displayed. Others may need location data to be visible to their employers during working hours, while they are on premise, but not outside of working hours or when they are off premise. Some users may want location data to be visible to their immediate families at all times. Others may want location data to be visible to all retail shops in the vicinity that offer deals of special interest.

Context can be loosely defined as the factors that may impact users' consideration regarding what is acceptable use of personal data. This can include a number of elements, such as the type of data, the nature of the interaction, how the data is used, and whether the use is something that is perceived to be of value to the user and/or the community, etc. In the physical world, people use a repertoire of personas as they interact with a variety of different people. The information they share depends on the persona they want to portray in each context. When these different personas are not recognised in the digital world, contextual integrity [8] is violated when the wrong data is shared in the wrong contexts. Sometimes users want to maintain integrity of contexts; sometimes they want to traverse contexts. What they choose will depend on their preferences, norms, culture and political and legal environment. These and other factors define a nuance of data use and a granularity of control that are not yet integrated into today's approach towards privacy and personal data management.

The lack of finer control may seem to be justified when one considers the overwhelming evidence that although users express a desire to control the use of their personal data, few actually do so. But there can be many reasons for not doing this – lack of simple and intuitive tools, lack of technical knowledge, lack of perceived value for doing so, too much information being presented, etc. Moreover, this lack of finer control can adversely impact the ability for data to flow, as discussed in Section 2 above.

There are multiple conceptual approaches to enabling simple but effective context-aware data use permission setting – not all of which require explicit control by the user. One simple approach is to enable users to set data use permission, not at the point of data collection, but immediately prior to when the data is used, thus giving the user a better sense of the benefit received. Recommender systems that rely on contextual and demographic data to provide personalised privacy settings provide another approach [9]. To accommodate changing norms, it is conceivable that "norms databases" could be established that capture how similar classes of data are used in different contexts. Such services could be provided by third-party consumer organisations similar to the services that consumers rely on today for ratings of everyday household products. Proxy services that learn user preferences and act on a user's behalf can also help facilitate context-aware data permission settings by users.

All these approaches would rely on the underlying data management and infrastructure layers to provide the mechanisms necessary to enable the specification and enforcement of such context-aware data use policies.

In the next sections, we discuss how, together, the elements in these layers enable different types of trust networks and enable new business models for CPDM.

4. Building a Context-Aware Data Ecosystem

4.1. The Need for Trust

As the volume of digital data generated increases and more businesses depend on data analytics to drive their business, people are growing increasingly concerned about their personal loss of control. This loss of control, and the asymmetry of power between individuals and businesses, are causing a crisis of trust. For CPDM, as defined above, to have a positive effect, it requires that the user be able to trust the various stakeholders in the data ecosystem to implement CPDM satisfactorily. As a result, she can trust that she can develop her social and private life as an individual. Let us therefore first examine the concept of "trust" in more details.

Trust is bestowed on an entity (whether a person or system) if that entity is perceived to be trustworthy in the given context and/or for the given task (see [8] and the contributions of Bus and O'Hara in [11] for more extensive discussions on trust). For example: an airline pilot and a surgeon would, in most cases, be considered trustworthy while performing their professional tasks. Most importantly, trust is strongly dependent on the context in which it is given. When applied in the digital environment, we often do not realise that many of the factors that play a role in trusting people or organisations offline are not available online (e.g. the experience of seeing a person's behaviour and presentation, hearing her talking, etc.) (see also [10]). Although it is clear to many that trust is not only derived from classical rational arguments, it might be good to note that in order to be trustworthy, systems must address two sets of elements in their design.

- Rational elements that provide concrete evidence to support trustworthiness. This includes ensuring mutual interest; providing insurance; being accountable and accepting liability for damage; complying with legal obligations and contractual conditions; providing technical assurances through Privacy by Design, Privacy Impact Assessment, quality of development; ensuring transparency, auditability and understandable information; being certified by trusted entities; ensuring good corporate practices with visible and enforceable codes of conduct; and in general engaging in behaviour that would ensure a good reputation. Note however, that this might be culturally determined.
- Emotional elements that appeal to users' perception of the entity. This includes building sympathy and a positive opinion, for example through being friendly or charming, providing consistent value, convenience and usefulness, an environment that feels familiar and friendly, wide social and/or cultural acceptance, or good customer support and maintenance services.

4.2. Organising Trust in the Digital Domain

Similar to the terminology confusion around identity, privacy and personal data management discussed in Chapter 1, we see a confusion in the various concepts of "trust" in connection to services, frameworks, networks, providers, etc. We will therefore first clarify the language we will use here.

In offline life, people (or more generally, stakeholders,[5] which can include enterprises, public services, institutions, etc.) generally work and live in communities that share some common beliefs and benefits. We indicated above that people tend to adopt different personae or identities in these communities, examples of which include families, states, sport clubs, healthcare systems, or financial networks. In such communities, participating stakeholders organise their social and economic transactions according to the rules and norms underpinning the trust that had been built up within the community. Given this, online CPDM and the digital ecosystem that facilitates it must have capabilities that can reflect the social structure of offline living.

We use the term *Trust Framework* (*TF*) for a set of rules and practices on PDM and digital transactions that aims to create trust between the stakeholders in one or more communities within a social structure. Figure 1 shows the components that can be included in a TF. The enforcement element in the TF ensures compliance with the rules specified. Entities engage with a TF via contracts that specify the set of relevant technical, business, and legal rules for the desired transactions. For example, a TF for healthcare in a country may specify the appropriate regulations, codes of conduct for data use and data sharing with others in the sector, the level of data security that must be supported, the business practices, and audit reporting. We call a TF a context-aware TF if it supports CPDM.

[5] A "stakeholder" in a community is used here to mean a natural person or legal entity that has an interest in cooperating with others in this community.

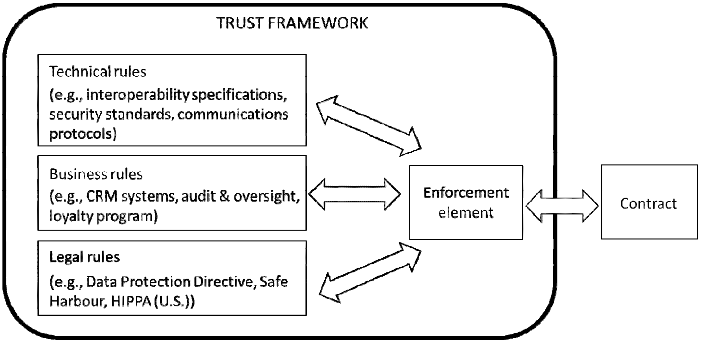
TRUST FRAMEWORK
Technical rules
(e.g., interoperability specifications, security standards, communications protocols)
Business rules
(e.g., CRM systems, audit & oversight, loyalty program)
Legal rules
(e.g., Data Protection Directive, Safe Harbour, HIPPA (U.S.))
Enforcement element
Contract

Figure 1. Definition of a trust framework.

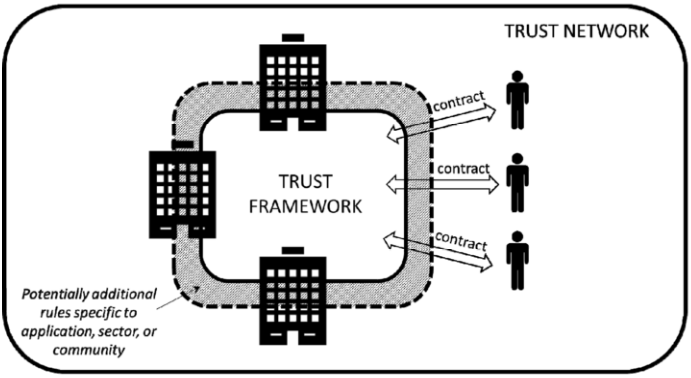
TRUST NETWORK
TRUST FRAMEWORK
contract
contract
contract
Potentially additional rules specific to application, sector, or community

Figure 2. Example of a service provider-centred trust network.

Finally, a *Trust Network* (*TN*) is a community of stakeholders that organises its digital transactions in accordance with a TF. As such, a TN can be considered an instantiation of a TF for a specific community. Figure 2 shows an example of a service provider-centred TN. In this TN, a group of service providers (SPs) have agreed to abide by a set of rules specified by a given TF. The TN also includes additional rules that may be specific to a given context, e.g. application, sector, or specific community. These additional rules may also be adopted for reasons of standardisation and interoperability. Users can enter into contractual relationships with one or more SPs in the TN if they trust the assurances that all SPs would abide by the rules specified by the TF as well as the additional rules included in this specific TN.

An example of a TN is a regional healthcare system, with hospitals, doctors, nurses and patients, that adopts a TF specified for healthcare. It might have additional rules regarding the use of personal data for research purposes. Other examples of TNs include bookshops and their customers; companies providing public transport services and their customers; banks and their customers; and also families or groups of friends communicating through social networks and other digital means. A TN which is an instantiation of a context-aware TF is called a context-aware TN.

We want to focus on context-aware TNs, and in particular those TNs that give stakeholders sufficient means to control their data. Moreover, we are particularly interested in TNs that ensure sufficiently high trust levels between the stakeholders regard-

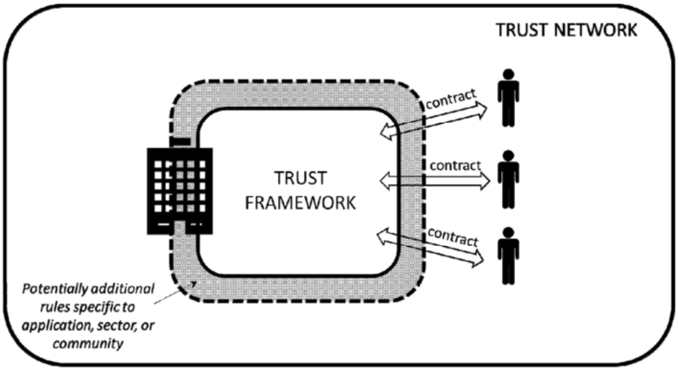
TRUST NETWORK
TRUST FRAMEWORK
contract
contract
contract
Potentially additional rules specific to application, sector, or community

Figure 3. Example of an O2M service provider-centred trust network.

ing regulatory compliance, context-aware PDM, and compliance with norms and rules and other service-provision conditions that are agreed between the stakeholders in the context-aware TN. Such a TN would need to implement appropriate services and applications at each of the layers discussed above in its implementation stack.

For some time now, service providers have been trying to build platforms that can support trust management in TNs. However, they often focus only on specific aspects (authentication, data security, reputation, etc.), leaving many other functions required for CPDM to be implemented by the TN itself. More recent developments have taken different and more promising directions. These are discussed in the following sections.

4.3. Trust Networks Involving Service Providers

A TN might use a digital platform from a third-party service provider to execute its transactions – such a service provider is termed a *TN Platform Provider* (*TNPP*).

A TNPP is considered one of the stakeholders in the TN and is responsible for ensuring PDM and service provisioning for the digital transactions in the TN in accordance with the TF, as well as any additional rules and conditions adopted in the TN. The platform provided by the TNPP often implements the services in the infrastructure and data management layers discussed above. It may or may not implement parts of the user interaction layer. Within the TN, "trust" is established when the TNPP can provide, through technology and contract, sufficient confidence about this to all other stakeholders in the TN.

There are two primary types of TN involving service providers.

1. *One2Many* (*O2M*): This TN is a transaction network between one service organisation and many end-users, supported by a TNPP. Figure 3 shows an example of an O2M service provider-centred TN. The service organisation can be a group of organisations if there is sufficient agreement between the members of such group to act together as a single organisation. The service organisation agrees, by contract with the end-users, how PD of the end-users in the TN can be used. Based on this contract, a technical interface can be implemented and managed by the TNPP, which is accountable to the TN by agreement that PDM and service management will follow contractual obligations. This could be achieved partly by technical means and partly through governance, depending on the agreement between the TN and the TNPP. The Dutch

start-up QIY [12] is an example of this type of TNPP. Examples of TNs that would leverage this type of TNPP include a large retailer with its loyalty customers, a large company with its employees, or a group of banks with their account holders.

The TNPP could bring a number of such TNs based on the same TF together on one platform. This would simplify governance, and allow standardisation and replicable implementation of (parts of) the interfaces. Note that one customer may also relate to more organisations on such a platform.

2. *Many2Many* (*M2M*): This TN reflects a community that may include service providers, individual users (e.g. customers, patients) and possibly other relevant organisations or institutions (e.g. consumer organisations, human right organisations, insurance companies, workers associations). Figure 2 and the discussion above describe an example of an M2M service provider-centred TN. The TNPP provides, and takes accountability for, the infrastructure that ensures compliancy of PDM and service transactions in the TN with the agreed rules and standards of trust. It may offer an API management portal for the SPs in the TN to act in a trustworthy way, and tools and services like "sticky policies", logging, auditing and random conformity tests, thus supporting trust as pre-agreed in the TN. Synergetics [13] is an example of such a TNPP.

Clearly both types of TNs create trust between their stakeholders, as they are all bound by the rules specified, including in the TF. The strength of that trust will depend on how the contracts were executed, the technology deployed by the platform selected to facilitate enforcement and auditability of the contracts, how the TN is governed etc. The choice of the type of TN to use depends on the service(s) and providers involved.

- If a large service organisation (e.g. a retailer or bank), for commercial reasons, does not want to cooperate with competitors, then O2M is likely to be a better solution.
- If a number of service providers in a given sector would like to cooperate to provide more comprehensive services to their customers (e.g. a healthcare network, a frequent-flier programme), an M2M would be more suitable.

In most cases, the M2M type is more efficient, as multiple service providers can share the same platform and functions for both the backend transactions and in the interactions with users. Despite their differences, the O2M and M2M TNs could use the same TF. Thus the rules implemented for trust in the PDM and service management could be the same.

4.4. User-Centred Trust Networks

The asymmetry of power between individuals and data-driven businesses that led to the emergence of unease and loss of trust has also led to the development of a specific type of TN, one which is focused on putting users in charge of the use of their data and motivating the development of new business models around personal data ecosystems. Variously termed vendor relationship management, personal clouds, personal data stores, and reputation systems, these ecosystems aim to correct the imbalance of power, and enable users to be in control of their data – including specifying what data can be accessed, how they can be used, who can use the data, and renegotiating policies for

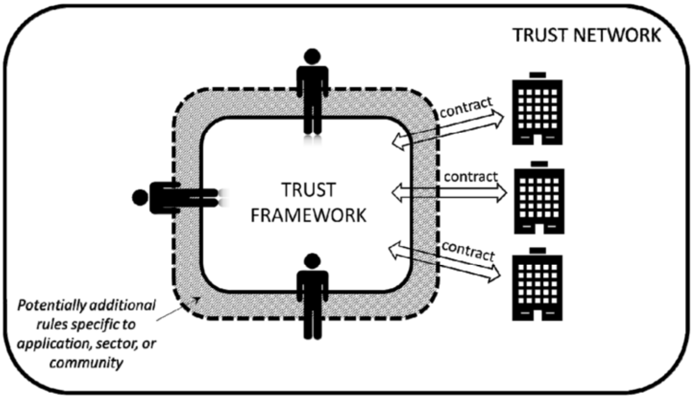
TRUST NETWORK
TRUST FRAMEWORK
contract
contract
contract
Potentially additional rules specific to application, sector, or community

Figure 4. Example of a user-centred trust network.

data use across all services with which they interact ([14]–[16] and the chapters by Heath et al. and Shadboldt (on midata) in this volume). The basic concept is a paradigm shift from businesses being in control of personal data to users being in control of personal data. In the former, businesses are collecting data to intuit user intent; in the latter, users, either individually or in trusted communities, come together to express a need that can then be satisfied by interested vendors. Figure 4 shows an example of an M2M user-centred TN where users come together to define a TF which specifies the policies under which their data can be used and service providers that agree to be part of the user-centred TN must abide by them contractually. Conceptually, there can be an O2M user-centred TN, where a single user would establish a TF that service providers would need to abide by if they want access to the user data. Realistically, this is an unlikely structure: individuals would be more likely to form a community around a TF rather than act on their own, as this would increase their bargaining power in attracting service providers to enter into contracts with them. Most of these TNs involve the use of a secured store for personal data at the infrastructure layer, and strict user-permission management at the data-management layer. An example of this is the "Life Management Platform" discussed elsewhere in this volume by Searls and Kuppinger.

The user-centred TNs and the service provider-centred TNs discussed above share common objectives in that they are both concerned with trustworthy enforcement of policies on data access and use. The difference lies in who defines the data policies and permissions and the rules applied in the TF: service providers or users. Current discussions of user-centred TNs require that vendors join the TNs (as shown in Fig. 4) since otherwise, users have no real control of the data use by service providers. Moreover, independent monitoring of the service provider actions would need to be addressed in this TN. This would require the service providers to cede control of data use to the users – a development that will probably require some time for widespread adoption. In a service provider-centred TN, the service provider binds himself to pre-described use of the data, and is being monitored and controlled on it by the TNPP and the governance structure. However, the contract to which users agree on entry to the TN presupposes informed choices about permitted data use – which might be an unrealistic assumption in many cases today.

Each of these classes of TN operating on its own would lead to suboptimal realisation of the value of data in the overall ecosystem. In practice, we see the two classes of

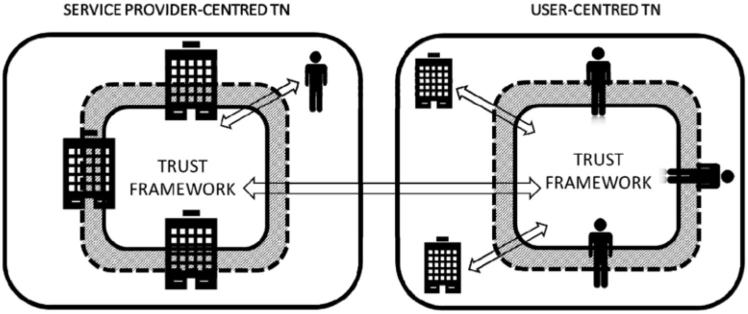

Figure 5. Co-existence of service provider-centred TN and user-centred TN.

TN co-existing, since they complement each other, and together can realise more value for all stakeholders in the ecosystem. Figure 5 shows this hybrid scenario, where a service provider-centred TN would negotiate with a user-centred TN for acceptable data policies, data permissions and operational rules.

4.5. Governance of TN

The essence of a TN is that the stakeholders can trust the network and the transactions – this cannot only be based on a one-time technology implementation, or on a contract signed by the TNPP. As such, governance is needed to ensure proper oversight, auditing, decision-making about members and monitoring procedures, and adapting the rules and conditions to changing circumstances. Governance will depend on culture, legislation and political environment. Transparency, checks and balances, and accountability might be relevant elements to consider. It seems clear that the TNPP would be subject to a representation of the community which forms the TN. Trust will be stronger if all stakeholders in the TN have sufficient representation in the control bodies. In addition, in many TNs it might be strongly advisable to ensure regular auditing by an independent party, or set the right of members to request investigation and reporting by independent experts in specific cases.

A TN should also be open to the use of various trust providers, e.g. identity or claim providers, personal data storage providers, reputation providers, etc. in order to stimulate competition, innovation and scalability.

4.6. Interoperability – Building the Ecosystem

Together, the various types of TFs, all kind of communities using TNs and other (context-aware) PDM services (including for identity, attribute and claim management) form a *data ecosystem*. This would reflect and integrate with offline life in society in a dynamic way, and grow and adapt with it.

Interoperability in the ecosystem between TNs can be greatly facilitated by using the same TF (i.e. the same trust principles/rules) and agreed-upon APIs. Certification of TNs and use of generally accepted standards and contracts will also support interoperability and the building of an ecosystem that allows users to switch easily from one environment (TN) to another; something which reflects the way people normally live.

Note that in an ecosystem of many TNs, it is desirable that an individual be able to manage the use of her personal data with contextually-consistent policies across the multiple TNs that she may belong to. This could be done within each TN, but would be more logically realised by a user-centred TN that can interoperate with the service provider-centred TNs that she currently belongs to. In the latter TN, she may want to specify policies for the use of her health data – for the health network that she belongs to – and they must be abided by and enforced; but for other personal data, she may want to leverage the commonly accepted data use policies that others in the user-centred TN have specified. In effect, the user-centred TN can act as a proxy for the user as appropriate, and furthermore, can reflect changing norms of the "user community" to the service providers. As explained above, TNs will operate within jurisdictions, societal norms and cultures, whilst existing power structures may also limit the choices of the user of the TN they want.

In addition to parsing the identity dimension in terms of the range of identities or personas a user can have (e.g. anonymous, pseudonymous, etc.), one can also consider the strength of the claims required. There are now alternative forms of verification, where instead of requiring a single verified claim, users are asked to provide information to answer a number of weak claims based on information that is available about them (e.g. last purchase, last location visited, first employer, etc.). Moreover, it is obvious that certain transactions do not need strong claims, and "Facebook-like" authentication suffices. Everything depends on the context, and the choices of the user and her knowledge about the consequences.

An ecosystem as described here could facilitate all three layers described in Chapter 3 above: the infrastructure layer, through the interoperable building blocks provided by TNs; the data management layer, through trust services for identity and claim management, auditing, compliancy enforcement, etc. and a first step towards user-centred context management layer through the various Trust Networks that allow definition of context and procedures that support it.

5. Interdisciplinary Dependencies

In Section 4 we have suggested a number of elements that, together, can lead to building a trusted PD ecosystem for our increasingly digital society. Such an ecosystem would be based on the implementation of CPDM for all members of this society. It is obvious that this needs to be a multi-disciplinary exercise. We can easily distinguish, but not disconnect, a number of perspectives or relevant disciplines that need consideration to achieve an overall acceptable and trustworthy CPDM.

Technical: we see strong trends towards trust networks, personal data ecosystems, policy-compliant data management, and personal data vaults, as described in Section 4. We also see "multiple identity" management and data minimisation through attribute-based verification and crypto-based credentials [17],[18]. Other issues relate to toolboxes for privacy by design, privacy impact assessment, or the broader social impact assessment. The role of technology in enabling alternative policy frameworks is only now starting to be examined.

Economic: discussions are ongoing on valorisation and propertisation of personal data. The current book is an example of this and will also contain many references. New and innovative business models are discussed based on different relationships between customer and service provider, as suggested above in the discussion of various types of TNs.

Legal: this relates to many issues including privacy as a human right, profiling, the newly proposed EC DP Regulation, law enforcement, and regulation in general. The relationship between technology and law is becoming more and more important, where technology may facilitate lawful processing, but may also lead to inscrutable systems imposing constraints which have no political or legal basis. Important in this context is the balance needed in the technology tools so as to allow an international framework that facilitates many different jurisdictions.

Socio-political: this includes the general trustworthiness of the infrastructure, institutions and applications that are the products of the above three aspects. Can we construct a digital world in which individual members can have confidence that they will find their place as respected members of the many online communities within and across jurisdictions? Some discussion on this can be found in the papers of Jacques Bus and Kieron O'Hara in [10].

Intergovernmental: the concepts of identity, PDM and privacy come from a physical world governed by sovereign states within national borders. Dealing with them in a "borderless" digital world requires international agreements and standards, and hence a debate that also considers the differences in appreciation of privacy and personal data management in the various global cultural blocs.

Any policy (public or private) will need to address these various aspects at the three levels mentioned above, and stimulate societal debate to develop policy that will both engender and strengthen CPDM, taking account of the perspectives we have discussed. The challenges to an international policy framework are enormous and the stakes high. The growth of international trade in the digital era will depend on it. The current work on an EU regulation for interoperability of authentication and digital signatures is only a first step.

The rules of governance of a TN, the establishment of electronic passports for international travel, and many international trade practices will be heavily affected by the development of a proper and sustainable personal data ecosystem.

It might be commendable to establish an international expert group, bringing together nations with the mission to establish guidelines for context-aware personal data management as a follow-up and extension to the OECD Guidelines on Privacy presented in 1980. In addition to experts representing the different perspectives discussed here, such group must also include representatives of user-centred and service provider-centred TN if a more optimal policy framework is to be developed.

6. Conclusions

The increasing integration of our on- and offline lives and the potential of big data are creating a pressing need for dialogues among stakeholders on a policy framework that can help realise the value of a data-driven ecosystem while also ensuring adequate rights and protection for individuals. A common language and reference model would greatly facilitate and promote such dialogues by providing consistency and context for these discussions.

In this paper, we have proposed a terminology framework that incorporates critical aspects of this conversation, including personal data management, context, and trust. We have also introduced a layered model to structure the dialogues on those elements that are required to develop ecosystems that can appropriately support context-aware personal data management. This is essential to the self-determination of individuals in the digital world, and the development of trustworthy user-centred data ecosystems that can enable this. We have also discussed new technologies and platforms that could support the development of such ecosystems.

Making this a reality will require unprecedented collaboration and cooperation between business, policy makers, and civil society, in addition to dialogues with individuals. These need to be multi-dimensional discussions that incorporate all aspects of the ecosystem: technical, scientific, economic, legal, and political. Moreover, this needs to be carried out on an international scale. These dialogues would be enhanced if there were a strong base of evidence that could inform the process and identify emerging issues along the way. And many dimensions will continue to evolve: technology, social norms, business models, and so on. Any policy frameworks must be able to accommodate this dynamic and evolving landscape.

The stakes are high. Tomorrow's economy *will* be data-driven. The discussions and recommendations we describe in this paper should be addressed urgently in countries that want to lead this global revolution and unleash the full potential of the digital economy.

List of Acronyms

CPDM	Context-aware Personal Data Management
PD	Personal Data
PDM	Personal Data Management
SP	Service Provider
TF	Trust Framework
TIP	Trusted Identity (or Claim) Provider
TN	Trust Network
TNPP	Trust Network Platform Provider

References

[1] World Economic Forum, Rethinking Personal data, Report 300512 (2012).

[2] James Q. Whitman, The Two Western Cultures of Privacy: Dignity versus Liberty, The Yale Law J. 113 (2004), 1152–1223.

[3] International Institute of Communications, Personal Data Management: The User's Perspective, London (2012).

[4] European Commission, Proposal for a Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data protection regulation), COM(2012) 11/4 draft.

[5] Paul Ohm, P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review 57 (2010), 1701; U of Colorado Law Legal Studies Research Paper No. 9–12.

[6] Ryan Singel, NetFlix Cancels Recommendation Contest After Privacy Lawsuit. Wired Magazine (2010, March 12).

[7] Arvind Narayanan, Vitaly Shmatikov, Robust De-anonymization of Large Sparse Datasets. In: SP '08 Proceedings of the 2008 IEEE Symposium on Security and Privacy. IEEE Computer Society (2008).

[8] Helen Nissenbaum, Privacy in Context – Technology, Policy and the Integrity of Social Life, Stanford (2010).

[9] Bart P. Knijnenburg, Alfed Kobsa, Making Decisios about Privacy: Information Disclosure in Context-Aware Recommender Systems. To appear in the ACM Transactions on Intelligent Interactive Systems (2013).

[10] Helen Nissenbaum, Securing Trust Online: Wisdom or Oxymoron? Working document (2001).

[11] Jacques Bus, Malcolm Crompton, Mireille Hildebrandt, George Metakides, Digital Enlightenment Yearbook 2012, IOS Press, Amsterdam (2012).
[12] QIY: see https://www.qiyfoundation.org/en/.
[13] Synergetics: see http://synergetics.be/.
[14] Doc Searls, The Intention Economy: When Customers Take Charge. Boston: Harvard Business Review (2012).
[15] Personal Clouds Wiki. (n.d.). Retrieved from http://personal-clouds.org/wiki/Main_Page.
[16] Respect Network. (n.d.). Retrieved from www.respectnetwork.com.
[17] ABC4Trust: see https://abc4trust.eu/.
[18] Kim Cameron, Reinhard Posch, Kai Rannenberg, Proposal for a Common Identity Framework: A User-centric Identity Meta-System. In: Kai Rannenberg et al. (Eds.), The Future of Identity in the Information Society, Springer (2009).
[19] Cynthia Dwork, Differential Privacy. In: Automata, Languages and Programming (pp. 1–12). Springer, Berlin Heidelberg (2006).